Environmental Research 135 (2014) 139–147

Contents lists available at ScienceDirect

Environmental Research

journal homepage: www.elsevier.com/locate/envres

Risk assessment's insensitive toxicity testing may cause it to fail

Vito A. Buonsante [a], Hans Muilerman [b], Tatiana Santos [c], Claire Robinson [d], Anthony C. Tweedale [e,*]

[a] ClientEarth, 36 Avenue de Tervueren, 1040 Brussels, Belgium
[b] Pesticide Action Network Europe, 1 Rue de la Pépinière, 1000 Brussels, Belgium
[c] European Environmental Bureau, 34 Boulevard de Waterloo, 1000 Brussels, Belgium
[d] Earth Open Source, 145-157 St. John Street, London EC1V 4PY, UK
[e] R.I.S.K. Consultancy, c/o EEB, 34 Boulevard de Waterloo, 1000 Brussels, Belgium

ARTICLE INFO

Article history:
Received 21 May 2014
Received in revised form
8 July 2014
Accepted 23 July 2014

Keywords:
Toxicity test methods
Test Guidelines-Good Laboratory Practices (TG-GLP)
Organization for Economic Cooperation & Development (OECD)
Risk assessment
Risk management

ABSTRACT

Background: Risk assessment of chemicals and other agents must be accurate to protect health. We analyse the determinants of a sensitive chronic toxicity study, risk assessment's most important test. Manufacturers originally generate data on the properties of a molecule, and if government approval is needed to market it, laws globally require toxicity data to be generated using Test Guidelines (TG), *i.e.* test methods of the Organisation for Economic Cooperation and Development (OECD), or their equivalent. TGs have advantages, but they test close-to-poisonous doses for chronic exposures and have other insensitivities, such as not testing disease latency. This and the fact that academic investigators will not be constrained by such artificial methods, created a *de facto* total ban of academia's diverse and sensitive toxicity tests from most risk assessment.

Objective: To start and sustain a dialogue between regulatory agencies and academic scientists (secondarily, industry and NGOs) whose goals would be to (1) agree on the determinants of accurate toxicity tests and (2) implement them (via the OECD).

Discussion: We analyse the quality of the data produced by these incompatible paradigms: regulatory and academic toxicology; analyse the criteria used to designate data quality in risk assessment; and discuss accurate chronic toxicity test methods.

Conclusion: There are abundant modern experimental methods (and rigorous epidemiology), and an existing systematic review system, to at long last allow academia's toxicity studies to be used in most risk assessments.

1. Introduction

Our objective in this article is to start an intensive dialogue between academic researchers and risk assessment agencies, on what are the determinants of reliable chronic toxicity test for a risk assessment of chemicals ('risk assessment'). Two opposing paradigms control toxicology – 'academic' and 'regulatory'. We define the former as investigations by researchers largely at universities and medical institutions. The latter however developed mostly in the nascent organic chemistry industry (especially synthetic pharmaceuticals), creating the toxicity test methods (Borzelleca, 1994) on which risk assessment relies on today, as we will demonstrate. We concentrate on the chronic exposure test, as it largely determines the regulation of agents in commerce, representing population-wide exposures. Risk assessment's methods were unified in a globally-adopted four-step paradigm by the US National Research Council's 'Red Book' (USNRC, 1983).

Other than an occasional regulator's generation of exposure data, a large information asymmetry exists in risk assessment. Companies investigate the physio-chemical character of molecules for marketable properties, including interactions with biologic systems. If a molecule appears worth commercialising, these data inform the necessary toxicity investigations (including on the agent's behaviour in organisms – adsorption to excretion), such as the dose level for *in vitro* and then in *vivo* acute toxicity tests. Such test results inform the dose levels for a sub-chronic exposure test, whose potency results finally informs the doses for the chronic toxicity test (Klaassen et al., 2013). This 'dose ranging' process is needed for a risk assessment, which aims to find a safe dose under all anticipated exposure scenarios.

The manufacturer performs these dose-ranging toxicity tests because the molecule promises profit if found safe enough to use.

* Corresponding author.
E-mail addresses: vbuonsante@clientearth.org (V.A. Buonsante), hans@pan-europe.info (H. Muilerman), tatiana.santos@eeb.org (T. Santos), claire.robinson@earthopensource.org (C. Robinson), ttweed@base.be (A.C. Tweedale).

http://dx.doi.org/10.1016/j.envres.2014.07.016

This basic conflict of knowledge and profit can cause a selective presentation of toxicity data when the agent undergoes approval for marketing (pre-market risk assessment). Indeed, many dozens of reviews show that findings of drug efficacy and risk are favourable to the manufacturer's interests (*e.g.* Sterne et al. (2008)), while publically funded tests return realistically mixed outcomes. Eight known such reviews exist for industrial chemicals (listed, Section 2.3), and all find the same correlation as the pharmaceutical reviews do.

1.1. How standardized toxicity test methods came to be dominate risk assessment

A key event in risk assessment occurred in the 1970s when a third of all the USA's regulatory chemical and pharmaceutical toxicity tests were suddenly brought into question by a whistle-blower who revealed massive fraud at just one laboratory used exclusively by industry, Industrial Bio-Test (Schneider, 1983). In response, the US Food and Drug Administration (US FDA) in 1978–1979 established mandatory Good Laboratory Practice (GLP) requirements for non-human tests (USFDA, 2014). GLP requires transparent, detailed documentation of the laboratory work and explicitly assigns responsibility for the various steps in an experiment, thereby increasing accountability; discouraging dishonest or criminal behaviour and enhancing the replicability (precision) of data.

US FDA's GLP was immediately adopted by the US Environmental Protection Agency (US EPA), then rapidly by regulatory agencies worldwide. The Organization for Economic and Commercial Development (OECD) member countries began adhering to GLP standards in their 1981 Mutual Acceptance of Data (MAD) decision (OECD, 2014a).

Crucially, MAD also marked the appearance of the OECD's Test Guidelines (TG) – standardized detailed protocols (methods) for performing toxicity tests. MAD requires that only TG and GLP-compliant toxicity tests be used in a risk assessment by any OECD member country. This strong OECD initiative – several detailed toxicity test methods begun and promulgated in just three years from the appearance of GLP – may indicate risk assessors' new determination to ensure reliable and standardized data. Equally, it may indicate industry's desire to retain control of the crucial data going into risk assessment. We speculate that after industry was forced to comply with GLP, it lobbied the OECD to use their existing (Borzelleca, 1994) insensitive toxicity test methods as *mandatory* TGs; in effect creating a global shield against use of academia's findings to determine risk.

We label the OECD's test methods 'TG-GLP', GLP being essentially a generic TG. Today, government bodies in developed countries oversee the creation/revision of toxicity test methods, all coordinated with the OECD's Working Group of the National Coordinators of the Test Guidelines Programme (WNT). The WNT is composed of the lead chemical agency of those countries (OECD, 2014b). The WNT accepts nominations for a new or revised toxicity test method from these national agencies, finalises it as a TG, then OECD promulgates it to member countries and lately especially to the rest of world (OECD, 2010). A few countries (such as the USA) create their own toxicity test methods, but these are entirely coordinated with TGs (USEPA, 2010).

Thus MAD drives statutorily-required use of TG-GLP in pre-market risk assessments across the world (OECD, 2014b; USEPA, 2014a); and, because many agents need a risk assessment in various jurisdictions over the decades, *the majority of all risk assessments are 'pre-marketing', and so must use TG-GLP methods* (parenthetically, many chemical uses require no approval to be used in commercial products, *e.g.* household cleaning or personal care products). We will show how the TG-GLP test methods, though with benefits, fail to detect much toxicity.

2. Discussion

2.1. How TG-GLP bars academia's studies from almost all risk assessment

Use of TG-GLP would provide academic investigators with adequate study power and some assurance of data quality. But science already has good data quality protocols, such as confidence intervals and peer review. The rather insensitive and artificial TG protocols hinder discovery. *Thus the net effect of requiring TG-GLP in risk assessments is to entirely exclude academia's results from most assessments.*

Regulatory agencies issue guidance on performing risk assessment, for use by their staff and industry (OECD, 2012a; USEPA, 1998; EFSA, 2010; ECbA, 2011). These reinforce the laws to use TG-GLP by advising that TG-GLP studies deliver the most reliable data for evaluating toxicity. A crucial underpinning to this conclusion is a published guide to data reliability authored by employees of the chemical multinational BASF (Klimisch et al., 1997). The guidances all say (*e.g.* ECbA, (2010)) that 'Klimisch' should be used to find the most reliable studies. But Klimisch simply states that TG-GLP studies return the most reliable data, giving them its top rank of '1' (it ranks other qualities, but 'reliability' is its key criterion).

The European Union (EU)'s Health Commissioner has testified: 'While it is correct that GLP does not evaluate the scientific quality and reliability of a study, it is the only internationally recognised quality system that monitors the organisational process and the conditions under which health and environmental safety studies are planned, performed, monitored, recorded, archived and reported.' (Dalli, 2011). Initial reviews suggest that adherence to TG-GLP criterion may not produce consistent results compared to other data quality criteria (Ågerstrand et al., 2014). Despite these concerns, the Klimisch criterion on the reliability of data is now an almost universally utilised (Ågerstrand et al., 2014) justification that TG-GLP methods produce the most reliable data.

Industry and regulators often say that academia's lack of dose ranging and heterogeneous methods make it impossible to evaluate the quality of their data; so academia's studies are only 'useful for generating hypotheses' (EFSA, 2012a, 2012b). They say that there is no barrier to use of academic studies, so long as they were of equivalent quality to TG-GLP studies...yet their criterion for study quality is TG-GLP (EFSA, 2012a; USNTP, 2013b).

Thus tens of thousands of published toxicity findings from academia are being ignored. At a conference of 300 senior risk assessors, not one when asked in plenary could spontaneously name a single pre-market risk assessment that did not rely on TG-GLP tests from industry to calculate its chronic safe dose (Tweedale, A.C., personal communication). Among the tens of thousands of pre-market risk assessments performed globally over several decades, we know of only one recent one (for vinyl cyclohexane, by the French agency ANSES – we would be interested to hear of any others).

Our intent is to break this circular logic, and get risk assessors to evaluate the reliability of academia's studies, not to simply exclude them.

The other type of chemical risk assessment is the post-marketing or 'review' risk assessment – often performed after the accuracy of a safe dose has been questioned. These are not required to, but often do, use a TG-GLP study as their key study. Nevertheless the accumulation of academia's published toxicity studies haltingly become relied on in post-market risk assessments. For example the US EPA's Integrated Risk Information

System (IRIS) (USEPA, 2014b) uses low-dose toxicity findings from academia to declare a safe chronic exposure dose for some chemicals. Industry greatly disputes the accuracy of IRIS's calculated safe doses (Rosenberg, 2011), so TG-GLP based LOAELs are only slowly replaced – *e.g.* the IRIS safe dose of bisPhenol-A (bPA) has not yet changed despite hundreds of published low dose toxicity findings. The US National Research Council thoroughly evaluated these controversies and concluded IRIS is very useful and becoming more so (with recommendations); while contradicting industry's chief claims (USNRC, 2014a).

2.2. The key sensitivity difference between academic and TG-GLP methods

TG-GLP methods have strengths. They make the parties of an experiment accountable for all decisions and test results, increasing data reliability. They are standardized and transparent, enabling inter-study comparison, and so are also easier to replicate (precision) than academia's studies. They use sufficient animal numbers to potentially detect weak effects, and they systematically test doses from poisonous to the end of poisoning. They are more specific (few false positives) than academia's tests.

But more specificity often *means less sensitivity*, a key criterion if health is to be protected. By using quasi-poisonous doses and other insensitivities (Section 2.4), the effects and toxicity thresholds elicited by chronic TG methods can be rather insensitive – liable to false negatives. Despite testing chronic exposure periods, the TGs detect largely the end of poisoning, as follows (Klaassen et al., 2013).

Chronic TGs aim to discover both a 'lowest observable adverse effect level' (LOAEL, the lowest dose an adverse effect is observed at), and a 'no observable adverse effect level' (NOAEL, the highest dose at which no adverse effect is observed); to set the safe daily dose for almost a lifetime of exposure. A 'Maximum Tolerated Dose' (MTD) is found by lowering the dose from acutely poisonous levels. That is, the chronic dose must be high enough to observe significant toxic effects from the limited number of animals that can be afforded by these expensive long term tests; but low enough to prevent most animals from wasting away from poisoning. Thus the doses of a TG-GLP chronic test typically range from the MTD (or just below it) for the highest dose, to 100-fold or so below the MTD for the lowest dose. In effect they are quasi-poisonous doses over most of a mammal's life.

Indeed, our readings of risk assessments show that the high doses of chronic TG elicit a slow poisoning: typically weight loss and gross histopathologic change of organs; usually the kidneys and liver, which try to excrete poisonous molecules. Poisoning by definition occurs in a linear dose–response fashion, so these high TG doses *crucially* tend to produce a LOAEL and a NOAEL, allowing a safe dose to be calculated *and the risk assessment to proceed* (if no LOAEL is found, the doses may be lowered and the test repeated).

2.3. But are these actual L/NOAELs?

No one disputes that finding a monotonic dose–response (D/R) relationship – no reversal of the dose–response (D/R) slope – supports causation, and they are also a common finding in academia's studies. *But monotonicity is not the only display of toxicity* and there are at least seven biologic reasons why a lower dose can be more potent than a higher one, indeed perhaps 20% of the time, finding over 800 examples (Vandenberg, 2014). Biochemistry – life – is resilient, but complex, and it often occurs at very low signal strengths (Ray and Gough, 2002); so life can be vulnerable to agents it did not evolve with.

But use of realistic doses are not even contemplated by the TGs; *e.g.* EU risk assessors recently dismissed non-monotonic teratogenic effects of the herbicide glyphosate, though they were found in tests designed and performed for risk assessment (Antoniou et al., 2012).

The insensitivity of the TG methods was directly demonstrated in a side-by-side comparison of test methods employed to investigate mammary gland toxicity (Makris, 2011), plus a companion paper of expert group analysis of the mammary gland microscope slides of the compared studies (Rudel et al., 2011); together they demonstrate that TG methods failed to detect various important signs of toxicity to this mammalian organ.

Support for low dose toxicity also comes from epidemiology's large, fast-growing literature, which correlates humanity's low-dose exposures with diseases, avoiding extrapolation across species. Epidemiologic methods are conservatively biased to the null hypothesis (Nachman et al., 2011): increasingly they are longitudinal (allowing cause to precede effect), use large sample sizes, have accurate exposure data, and control confounders better (Nachman et al., 2011). In sum, epidemiology increasingly contributes to establishing causation. The US EPA's 'Dioxin Reassessment' (USEPA 2010), the most extensive risk assessment ever – ongoing for more than 25 years – relies on epidemiology to find (as draft) that very low dioxin doses are dangerous.

Finally, the insensitivity of the TG test methods is evidenced by at least eight published reviews comparing industry's toxicity

Table 1
In-vivo refutations of risk assessment chronic toxicity L/NOAELs.

Chemical	L/NOAEL in a risk assessment	Potency in published literature
Glyphosate, herbicide	100 mg/kg d- (*circa* NOAEL of industry's key studies)	4.87 mg/kg d-, no NOAEL, tests glyphosate (Benedetti et al., 2004).
HexaBromo CycloDodecane (HBCDD) Flame Retardant	100 ppm (8.1–21.3 mg/kg d-) NOAEL REACh Authorisation (Saegusa et al., 2009).	0.9 mg/kg d- single oral dose altered mouse spontaneous behaviour, no NOAEL Eriksson et al., 2006).
Tri-n-Butyl Tin Molluscide, fungicide.	25 μg/kg d- NOAEL (Vos et al., 1990).	0.5 μg/kg d- (5.42 nM in water, *circa* human body burdens): increased obesity parameters, through F3 unexposed generation (Chamorro-Garcia et al., 2013).
2,4-D, herbicide	62.5 mg/kg d- LOAEL (Charles et al., 1996).	2.5 mg/kg d- (in food; also as single i.p. dose): hormone alteration, lactation problems (Stürtz et al., 2010).
Cadmium	10 μg/kg d- via food NOAEL (USEPA 2014c).	5 μg/kg (~27 nM/kg) single i.p. inj. proliferated F reproductive organs in maturity, no NOAEL (Johnson et al., 2003).
Arsenic	170 μg/L (drinking water) LOAEL (USEPA 2014d)	10 μg/L (10 ppb, EPA alleged safe level) in *ad libid.* water: decreased growth *in utero* and F1, no NOAEL (Kozul-Horvath et al., 2012).
Formaldehyde	82 mg/kg/d-oral LOAEL/15 mg kg d- NOAEL: decreased rat organ/body weights, (USEPA, 2014e; ECbA, 2014); 3.2 mg/m³ chronic inhalation NOEC (ECbA, 2014); 0.1 mg/m³ chronic *local* eff. NOEC (ECbA, 2014).	~1 mg/kg d- (10 mg formaldehyde/L water *ad libid.* S-D rats): cancers, no NOAEL (Soffritti et al., 2002). 0.1 mg/m³: asthma (human), no NOAEL (Casset et al., 2006); 0.52 mg/m³ asthma, no NOAEL (Qiao et al., 2009).

Not just endocrine disruptors cause low-dose toxicities.

studies with academia's studies (Bekelman et al., 2003; Diels et al., 2011; Domingo and Bordonaba, 2011; Fagin and Lavelle, 2002; Hayes, 2004; Lesser et al., 2007; Swaen and Meijers, 1988; vom Saal and Hughes, 2005). These clearly show that industry studies find little or no toxicity, while publicly funded studies of the same chemical realistically yield mixed results, including many findings of low dose toxicity.

Agents including arsenic, lead, mercury, ozone, particulate matter and dioxin-like compounds have had their 'safe' dose repeatedly lowered over the decades until it is generally conceded that they may have no safe exposure level, but this has not occurred for any strictly commercial agents. Rather, regulatory agencies while respecting the findings of academic science as 'hypothesis-raising', seem to require the more unrealistic standardized data resulting from dose-ranging.

Yet one may pick any well-known agent (so it has a large enough published toxicity literature) and see published findings of chronic mammalian toxicity at doses lower than the LOAEL claimed in its risk assessments. We list a few dozen examples in Tables 1 and 2 (some are even more potent than some of the LOAELs in the IRIS database). Parenthetic to our purpose, we note that ecological risk assessment seldom performs any chronic exposure test at all. However fixing the mammalian chronic assay's insensitivity would benefit all species.

2.4. What makes TG-GLP tests so insensitive?

Any study has shortcomings, but why do TG-GLP methods so regularly fail to find toxicity at the levels that organisms are typically exposed to, when other test methods do? Here are the main insensitivities of TG chronic test methods:

(1) A TG test sacrifices the animals at the end of dosing, at human equivalent of circa 60 years old, before most chronic disease manifests – *e.g.* 77% of malignant tumours are diagnosed after age 55 in the USA (ACS, 2013).
(2) Not enough tests of developmental toxicity are done, despite the complex vulnerability of development, which drives much disease, even in adulthood (Hanson and Gluckman, 2011).
(3) Data from concurrent negative controls in a TG-GLP experiment are allowed (OECD, 2012b) to be diluted, even overridden, by historical control data drawn from experiments carried out in a wide range of different conditions (accordingly, they are used in many risk assessments). Some of the variables not well controlled when using the often secret historical controls include strain and origin of animals, laboratory in which the experiment was carried out, dietary factors; environmental contaminants in air, bedding, food, and water; differences in diagnostic criteria among pathologists, and the year in which the experiment was performed; all which can produce very different results (Haseman, 1984; Hardisty, 1985).
(4) Positive controls (when feasible) limit false negative results (Myers et al., 2009), but are never mentioned in the TGs or in guidance.
(5) Toxicity is almost always detected with the light microscope and a few gross biochemistry measures, rather than also employing academia's advanced imaging and biochemistry methods (Koshland Jr., 1998).
(6) As just described, the TG's high dose levels tend to elicit a quasi-poisoning syndrome that is irrelevant to the effect of the doses encountered in the biosphere, which remain untested by TGs.

Table 2
In-vivo refutations of most protective TTC's assumed NOAEL, 150 μg/kg bw d-.

Chemical	N/LOAEL (μg/kg bw d-)	Times < TTC NOAEL (×)	Reference
Diethylstilbesterol (DES)	0.018	8333	Bøgh et al. (2001)
Bisphenol-A	0.025	6000	Muñoz-de-Toro et al. (2005)
HCB + 1,2,3-TCBenzene	0.1	1500	Vaikusz et al. (2011)
BDE-47	0.2	750	Abdelouahab et al. (2009)
Ethinylestradiol (EE2)	0.2	750	Vosges et al. (2008)
TriButylTin	0.4	375	Meador et al. (2011)
Dicamba	0.9	167	Cavieres et al. (2002)
Atrazine	1	150	Belloni et al. (2007)
Bisphenol-A	2	150	Melnick et al. (2002)
Fenarimol (pyrimidine fungicide)	2	75	Park et al. (2011)
BDE-47 (Br diphenyl ether)	2	75	Suvorov and Takser (2011)
Deltamethrin	3	50	Issam et al. (2009)
Dieldrin	5	30	Walker et al. (1969)
Haloxyfop methyl	5	30	USEPA (2014f)
Triflumazole	8.6	17	Li et al. (2012)
Di-n-butyl phthalate	10	15	Hoshi and Ohtsuka (2009)
Perchlorate	10	15	Yu et al. (2002)
PFOA	10	15	Macon et al. (2011)
Octylphenol	10	15	Alworth et al. (2002)
Methoxychlor	10	15	Bøgh et al. (2001)
DEHP (phthalate)	15	10	Andrade et al. (2006)
o,p'-DDT	18	8	Palanza et al. (1999)
Methoxychlor (two at same dose)	20	8	Cioiosa et al., (2007), Armenti et al. (2008)
Toxaphene	50	3	Olson et al. (1980)
BDE-99	60	$2\frac{1}{2}$	Kuriyama et al. (2007)
Nonylphenol	100	$1\frac{1}{2}$	Yu et al. (2011)

For many years industry has promoted the Threshold of Toxicologic Concern (TTC) as a substitute for chronic toxicity testing. A TTC is a claimed safe dose for effect categories of agents (genotoxic, endocrine-disrupting, etc. (the latter's appropriateness for the TTC is still being debated). A TTC is set below the LOAELs of up to a few hundred existing toxicity results; but with the usual preference for TG-GLP-generated results. Consequently, it is just as easy to find examples of more potent toxicity than a TTC as it is for those in Table 1. *Here even the most protective of the TTCs*, for Cramer Class III agents: **1.5 μg/kg d-** is shown to not be protective. Many of our examples dose by feed/gavage, so the elicited toxicity is after first pass metabolism and excretion. Our example doses are mostly LOAELs while the TTC uses mostly NOAELs, so our refutations are stronger yet. We include examples of natural hormones to emphasise how potent hormones can be. Finally we assume the standard 100-fold uncertainty factors went into this TTC, for a putative 'universal' NOAEL of 150 μg/kg d-.

In contrast, academic researchers develop whatever methods suffice to investigate an agent's chronic toxicity; and they use the time-tested quality control methods of peer review and publishing (albeit these are undergoing challenge). The later allows further peer critique. Also, a study's methods and raw data are available for inspection after publication, according to science customs.

The US National Toxicology Program's publically funded chronic toxicity tests also – as TG-GLP methods do – use doses high enough to reliably detect effects for a set number of test animals. Nonetheless their results have been tested and found to predict carcinogenicity (Huff, 2002; Maronpot et al., 2004), and continually improve the sensitivity of their methods. The Ramazzini Institute near Bologna, Italy employs an opposite approach, using as many animals as are needed to reliably detect low dose chronic effects. Their tests regularly find toxicity at doses deemed safe under TG-GLP methods (Chiozzotto et al., 2012; Soffritti et al., 2008) and their data were recently partially validated and recommended for use in risk assessment (Box 1). Note that in the USA and likely elsewhere, animal welfare concern prevents federally-funded life science academics from using larger animal groups than needed to detect a significant effect (vom Saal and Hunt, 2012); thus TG-GLP proponents can misleadingly claim that academia's studies are too underpowered compared to theirs; as the USFDA did in a 2012 post-market assessment of bPA risks. Yet proper controls and other method issues have a greater influence than group size does on sensitivity (A. Soto, personal communication).

Regulators correctly note (USNTP, 2013a) the increasing use of the Benchmark Dose (BMD) to establish a risk assessment's safe doses. Rather than searching for a LOAEL which assumes no toxicity is possible below it, a BMD is the dose at which toxicity first manifests. This encourages testing of low doses (though distinguishing harmful from harmless changes may be controversial). A validation of BMDs using 352 long studied chemicals not only verified that more testing at low doses improved the accuracy of a dose/response data set, but that these lower doses frequently caused toxicity below the alleged NOAELs (Wignall et al., 2014). Allowing academia's low dose toxicity tests to be considered in risk assessment would encourage wider adoption of BMD in risk assessment.

2.5. *Persuading risk assessors to consider academia's toxicity data*

To recapitulate, academia's toxicity studies are excluded from most risk assessments, which instead use data from the somewhat artificial and insensitive TG-GLP tests. Rapidly gaining reliability and realism are *in vitro* and *in silico* test methods (Birnbaum, 2013), as well as substitutions of models for toxicity testing, *e.g.* the Threshold of Toxicologic Concern (TTC – see Table 2 for description), which regularly are proposed to improve risk assessment. However, the chronic mammalian bioassay should be the main focus of improving accuracy – as it most realistically models human risks.

Given the demonstrated insensitivities of the TG-GLP methods, there is an urgent need for national chemical safety regulators to dialogue with academic researchers, to intensively debate the determinants of accurate (both sensitive and specific) and precise (replicable) chronic toxicity test methods – *i.e.*, of reliable data. The following modifications to pre-market risk assessment are indicated. They are long-term goals, as achieving them will require much dialogue.

2.5.1. *For agents not previously assessed: independent testing*

As described, the inventor initially has all knowledge on their agent. Their role in the future should be only to provide the agent and their data on it (with confidentiality of competition-sensitive business information); and to pay the cost of independent testing (only through fees paid to national treasuries, in order to dilute their influence). Financially independent academia should be statutorily declared to be the rebutably-preferred source of data

Box 1–Further ways to improve risk assessment.

Adopt methods and the offer of learned academic societies

An offer of learned societies of the life sciences (ASHG, 2011) to lend their unmatched expertise in investigating toxicity must be seriously considered by regulators. Such a sensitive chronic toxicity protocol is already in use at laboratories such as Italy's Ramazzini Institute (at least one academic lab in the USA, perhaps elsewhere, are doing the same). Despite the expense, the Ramazzini laboratory estimate human exposure levels to determine the dose and thus the necessary animal groups' size. They expose animals *in utero* and through development and allow test animals to live out their lives – at least 120–130 weeks of age for rodents – as chronic diseases take time to develop (Chiozzotto et al., 2012). Their 'GLP Life Test' laboratory is GLP-certified, a key demand of risk assessors. While regulators have cited false positive cancer slide readings (infections, not cancer) by Ramazzini Institute, a new leading experts examination of their microscope slides prove that any confounding by infections is limited to three cancer types. Otherwise, their conclusion is that Ramazzini's sensitive test methods are especially useful for risk assessors (Gift et al., 2013) – which USEPA and the EU's EFSA had previously rejected.

Adopt NIEHS's TiPED

Recent federal USA initiatives on risk assessment (Birnbaum et al., 2013) include a Tiered Protocol for Endocrine Disruption (TiPED) framework by the US National Institute of Environmental Health Sciences (NIEHS) for testing the effects of low doses (Schug et al., 2013). TiPED is an ideal risk assessment framework for all endpoints, not just endocrine effects. TiPED would integrate all available data on an agent, from modelling through to chronic mammalian exposures, with all methods to be kept at the 'cutting-edge' by the best scientists in these various fields.

Rescue NRC's Silver Book recommendations

We strongly support the 'Silver Book' recommendations of the US National Research Council on re-inventing risk assessment (USNAS, 2009), which inter alia would expand use of low dose test methods by abandoning the assumption of a threshold (safe) dose (only carcinogenicity tests currently do). But the implementation of these recommendations seems to have been entrusted to the very parties – industry and regulators – who believe today's insensitive TG-GLP test methods are superior; with very heavy involvement by industry (ARA, 2014). While participants in this 'Alliance for Risk Assessment' (ARA) project are aware that more sensitive toxicity methods than the TG-GLP exist (ARA, 2013), most of those involved appear to believe that improvements to the existing methods – *e.g.* more data on mode of action or exposures – will make risk assessment 'fit for purpose', the Silver Book's rubric for better risk assessment. Not only are those improvements not needed to find toxic effects, but ARA is promoting (ARA, 2013) methods such as the industry-promoted TTC, which abandon any toxicity testing at all (Pesticide Action Network Europe, 2012).

in risk assessment. Independent academics could then be contracted by governments to test the agent and analyse the inventor's data – from its physio-chemical properties through to any toxicity tests provided. When data is conflicting or lacking, there should be a statutory precautionary bias when risk managers decide the fate of an agent (along with further tests to decide the question).

2.5.2. For previously-assesed agents: critical (systematic) review

A risk assessment's first step, a literature review, is critical. Yet these are pre-judged by the requirement for TG-GLP, usually via Klimisch, to summarily dismiss other findings.

Mandates on industry to evaluate all literature on an agent are starting to appear (*e.g.* in the EU's chemicals (REACh) and pesticide laws), but our audits (ClientEarth, 2013; EEB, 2012, and one due Sept. 2014 by Pesticide Action Network Europe) show these mandate so far elicit reporting of no more than a quarter of academia's published findings on an agent, with some companies failing to report *any* published study. This attempt to improve the critical first step of risk assessment is failing. Critical reviews on an agent *are* regularly published by academic investigators – *e.g.* on bPA (Richter et al., 2007) – which would be a useful starting point for a risk assessment, if an up to date one exists.

But to systematically determine the most reliable data, risk assessors should adapt the 'critical (systematic) review' methods of 'evidence based medicine' – the result of clinician's struggles to interpret conflicting medical findings – chiefly the Cochrane Collaboration (2014). A critical review aims 'to minimise bias by using explicit, systematic methods' to review all the literature, then critique it with objective, evidence-based criteria (Green et al., 2008), creating the most reliable synthesis of current knowledge (Woodruff and Sutton, 2010). Transparent presentation facilitates these difficult evaluations, typically reaching consensus (Evans et al., 2011).

Risk assessment agencies are moving towards more systematic reviews, *e.g.* the Navigation Guide (Woodruff and Sutton, 2011) and USNTP's Office of Health Assessment and Translation, OHAT (Bucher et al., 2011). Journal editors are beginning to screen animal experiment manuscripts with the ARRIVE criteria to improve their reproducibility (Tilson and Schroeder, 2013). Even traditional risk assessment agencies and industry are moving towards systematic review, *e.g.* the EU's Food and Feed Safety Assessments (EFSA, 2010); the Evidence-Based Toxicology Collaboration (Hartung, 2009).

Criteria on data quality are the key to successful critical/systematic review. Some elements of TG-GLP tests, *e.g.* transparency of reporting, score high. Yet *there is evidence that non TG-GLP methods (including limiting financial conflicts, as the Cochrane guidelines have proposed) create more reproducible results* (Vesterinen et al., 2013). And a review of systematic review's criteria finds that just one of 30 had been well validated (tested), and most appear to promote insensitive toxicity test methods such as the TG-GLP (Krauth et al., 2013). Another such comparison also supports use of rigorous systematic review criteria, far beyond what TG-GLP test methods offer (it tested 12 published bPA chronic toxicity studies against a rigorous set of method criteria, and even most of the reviewed authors agreed with their criticisms) (Ågerstrand et al., 2014).

2.5.3. How to modify risk assessment: summary

Significant data gaps are found in all risk assessment; they should be filled using the above procedure to test new agents. Risk assessment would then proceed as today: use the most potent of the validated chronic NOAELs or LOAELs (or BMD) to base a safe dose for all anticipated exposures.

Academic researchers could greatly contribute by using the useful attributes of the TG-GLP methods; especially to homogenise their toxicity test methods (to increase the comparability of results) as far as possible, without sacrificing their freedom to hypothesise and test. Regulators making risk assessment and management decisions should specify exactly their data needs, in the following dialogue which we propose.

Industry would defend its interests in this system by providing data showing that potent toxicity findings are false positives – indications of test methods that overly sensitive and not specific enough – although as with anyone's data, their findings would be subject to independent confirmation.

Industry has greatly increased its funding of academia in recent decades (Zinner et al., 2009), raising doubts about the reliability of academia's research. But academic researchers are historically independent-minded, and academia's on-going publication of so many low dose toxicity findings seems to show that we can rely on them. Journals are greatly improving disclosure of financially conflicting interests (CoI), and everyone should speak up against threats to academic objectivity. A role-playing study showed that disclosure reduces both the number of CoI and how biased expert advice is (Sah and Loewenstein, 2014).

Specific initiatives to make risk assessments more sensitive are presented in Box 1.

2.6. Starting a dialogue

We do not expect the massive global system for assessing the risks of chemical and other agents to change rapidly. Rather we aim to expand a dialogue that recently began between representatives of the opposing toxicology paradigms we have described. It erupted with an editorial from the editors-in-chief of 18 traditional toxicology journals (and 71 supporting researchers), saying that the European Commission should make no changes to risk assessment to assess endocrine disruptor risks, especially it should keep assuming a safe dose exists (Dietrich et al., 2013), to avoid the EU list for ban or restrictions. That elicited two ripostes from an even greater number of editors and supporting scientists: in Environmental Health (Bergman et al., 2013a) and in Endocrinology (Gore et al., 2013). Importantly, the journalists at Environmental Health Network revealed that of the 18 authors of Dietrich 2013, 17 failed to disclose financial ties to industries whose agents are subject to risk assessment, as did at least 40 of the 71 supporting scientists (EHN, 2013). But at least a dialogue on what is reliable toxicity data has begun.

In addition, the NIEHS organised a 2012 global workshop in Berlin (NIEHS, 2012) whose purpose was for regulators to talk with the academic researchers who frequently find non-monotonic and low dose toxicities; aiming to incorporate such results into risk assessment. But the challenge that non-monotonic results pose to classic regulatory toxicology's core paradigm, '*the dose makes the poison*,' is hard to exaggerate. The NIEHS is encouraging continuation of this dialogue. Helpfully the US National Academies of Sciences has advised USEPA to re-consider the evidence of low dose toxicity (namely non-monotonic), and to better adapt risk assessment to account for non-monotonic risks (USNRC, 2014b).

3. Conclusion

The immediate task for risk assessment's stakeholders is to develop the dialogue on the accuracy of TG-GLP versus academia's test methods. *The OECD's WNT committee a natural forum for further dialogue.* They originate and revise the TGs that are in global use, and WNT members are representatives of the largest national chemical agencies. They already discuss *ad hoc* test method issues with

stakeholders, including some academic researchers (OECD, 2014b). However, modern endocrinologists and other academics have the knowledge the WNT needs to turn the TGs into more sensitive toxicity methods – but their methods are not used. Anyone interested in this dialogue could contact us, *inter alia*.

People become upset (EC, 2013) when told they are permanently contaminated (USCDC, 2013) with synthetic molecules they did not evolve with. They pay taxes for academics to research those risks with the best techniques, so their trust in regulators erodes when they discover that this high quality data is of no interest to the regulators, who instead use data from the party whose interests conflict with knowledge. Toxicity testing also comes at great cost to animal welfare – all the more reason that its results be reliable, reflecting all methods validated in a scientific field (*i.e.*, with accurate data, duplicate animal testing is reduced). Many non-communicable chronic diseases are increasing in incidence, chemicals being a leading suspect (Bergman et al., 2013b). The primary prevention – not treatment or adaption – of any such calamity cannot occur without more sensitive toxicity test methods.

Funding sources and competing interests

No funding was used to write this article. The authors all advocate for a healthier environment via their non-profit organisations; thus any financial gains to them or their employers from publication of this paper will be *de minimus* (membership increases, etc.). We have no other conflicting interests.

Acknowledgments

We are very grateful to Paul Whaley of The Cancer Prevention and Education Society, UK, for his analysis of systematic (critical) review.

References

ACS (American Cancer Society), 2013. Cancer Facts & Figures Homepage. Available: ⟨http://www.cancer.org/research/cancerfactsstatistics/index⟩ (accessed 30.04.14).

ARA (Alliance for Risk Assessment), 2013. Beyond Science and Decisions, From Problem Formulation to Dose-Response Assessment Workshop, 6 May 2013. Available: ⟨http://www.allianceforrisk.org/ARA_Dose-Response.htm⟩ (accessed 30.04.14).

ARA (Alliance for Risk Assessment), 2014. Beyond Science and Decisions. Sponsors Homepage. Available: ⟨http://www.allianceforrisk.org/ARA_Dose-Response_Sponsors.htm⟩ (accessed 30.04.14).

ASHG (American Society of Human Genetics), American Society for Reproductive Medicine, Endocrine Society, Genetics Society of America, Society for Developmental Biology, Society for Pediatric Urology, et al., 2011. Assessing chemical risk. Societies offer expertise. Science 331 (6021), 136.

Abdelouahab, N., Suvorov, A., Pasquier, J.C., Langlois, M.F., Praud, J.P., Takser, L., 2009. Thyroid disruption by low-dose BDE-47 in prenatally exposed lambs. Neonatology 96, 120–124.

Ågerstrand, M., Edvardsson, L., Rudén, C., 2014. Bad reporting or bad science? Systematic data evaluation as a means to improve the use of peer-reviewed studies in risk assessments of chemicals. Hum. Ecol. Risk Assess. 20, 1427–1445.

Aiworth, L.C., Howdeshell, K.L., Ruhlen, R.L., Day, J.K., Lubahn, D.B., Huang, T.H., et al., 2002. Uterine responsiveness to estradiol and DNA methylation are altered by fetal exposure to diethylstilbestrol and methoxychlor in CD-1 mice. Effects of low versus high doses. Toxicol. Appl. Pharmacol. 183, 10–22.

Andrade, A.J., Grande, S.W., Talsness, C.E., Gericke, C., Grote, K., Golombiewski, A., Sterner-Kock, A., Chahoud, I., 2006. A dose response study following in utero and lactational exposure to di-(2-ethylhexyl) phthalate (DEHP): reproductive effects on adult male offspring rats. Toxicology 228, 85–97.

Antoniou, M., Habib, M., Howard, C.V., Jennings, R.C., Leifert, C., Nodari, R.O., et al., 2012. Teratogenic effects of glyphosate-based herbicides: divergence of regulatory decisions from scientific evidence. J. Environ. Anal. Toxicol. S4, 006.

Arnenti, A.E., Zanna, A.M., Passantino, L., Uzunoca, M., 2008. Developmental methoxychlor exposure affects multiple reproductive parameters and ovarian folliculogenesis and gene expression in adult rats. Toxicol. Appl. Pharmacol. 233, 286–926.

Belloni, V., Alleva, E., Dessi-Fulgheri, F., Zaccaroni, M., Santucci, D., 2007. Effects of low doses of atrazine on the neurobehavioural development of mice. Ethol. Ecol. Evol. 19, 309–322.

Benedetti, A.L., Vituri, C. de L., Trentin, A.G., Domingues, M.A., Alvarez-Silva, M., 2004. The effects of sub-chronic exposure of Wistar rats to the herbicide glyphosate-biocarb. Toxicol. Lett. 153, 227–232.

Bergman, Å., Andersson, A.M., Becher, B., van den Berg, M., Blumberg, B., Bjerregaard, B., et al., 2013a. Science and policy on endocrine disrupters must not be mixed. a reply to a "common sense" intervention by toxicology journal editors. Environ. Health 12, 69.

Bergman, Å., Heindel, J.J., Jobling, S., Kidd, K.A., Zoeller, R.T., 2013b. State of the Science of Endocrine Disrupting Chemicals. In: United Nations Environment Programme, World Health Organization (Eds.), WHO, Geneva, Switzerland (296 pp.).

Birnbaum, L.S., 2013. 15 Years out. Reinventing ICCVAM. Environ. Health Perspect. 121, a40.

Birnbaum, L.S., Thayer, K.A., Bucher, J.R., Wolfe, M.S., 2013. Implementing systematic review at the national toxicology program. Status and next steps. Environ. Health Perspect. 121, a108–a109.

Borzelleca, J., 1994. History of toxicology. In: Hayes, W. (Ed.), Principles & Methods of Toxicology. Raven Press, New York (1468 pp.).

Bucher, J.R., Thayer, K., Birnbaum, L.S., 2011. The Office of Health Assessment and Translation. A problem-solving resource for the National Toxicology Program. Environ. Health Perspect. 119, a196–a197.

Bøgh, I.B., Christensen, P., Dantzer, V., Groot, M., Thøfner, I.C.N., Rasmussen, K.K., Schmidt, M., Greve, T., 2001. Endocrine disrupting compounds. Effect of octylphenol on reproduction over three generations. Theriogenology 55, 131–150.

Bekelman, J.E., Li, Yan, Gross, C.P., 2003. Scope and impact of financial conflicts of interest in biomedical research. J. Am. Med. Assoc. 289, 454–465.

Casset, A., Marchand, C., Purohit, A., le Calve, S., Uring-Lambert, B., Donnay, C., et al., 2006. Inhaled formaldehyde exposure. Effect on bronchial response to mite allergen in sensitized asthma patients. Allergy 61, 1344–1350.

Cavieres, M.F., Jaeger, J., Porter, W., 2002. Developmental toxicity of a commercial herbicide mixture in mice. I. Effects on embryo implantation and litter size. Environ. Health Perspect. 110, 1081–1085.

Chamorro-Garcia, R., Sahu, M., Abbey, R.J., Laude, L., Pham, N., Blumberg, B., 2013. Transgenerational inheritance of increased fat depot size, stem cell reprogramming, and hepatic steatosis elicited by prenatal exposure to the obesogen tributyltin in mice. Environ. Health Perspect. 121, 359–366.

Charles, J.M., Bond, D.M., Jeffries, T.K., Yano, B.L., Stott, W.T., Johnson, K.A., et al., 1996. Chronic dietary toxicity/oncogenicity studies on 2,4-dichlorophenoxyacetic acid in rodents. Fundam. Appl. Toxicol. 33, 166–172.

Chiozzotto, D., Panzacchi, E., Tibaldi, M., Laurirola, M., Belpoggi, F., 2012. Prenatal exposure increases risk of cancer. the experience of the Ramazzini Institute [Poster]. In: Presented at 3rd Childhood Cancer Conference, London, 24–26 April 2012. Available: ⟨http://www.childhoodcancer2012.org.uk/abstracts.asp⟩ (accessed 30.04.14).

ClientEarth, 2013. REACH registration and endocrine disrupting chemicals. Available: ⟨http://www.clientearth.org/reports/reach-registration-and-endocrine-disrupting-chemicals.pdf⟩ (accessed 18.01.14).

Cochrane Collaboration, 2014. The Cochrane Collaboration Homepage. Available: ⟨http://www.cochrane.org⟩ (accessed 30.04.14).

Dalli, J., 2011. Answer given by Mr. Dalli on behalf of the Commission. European Parliament Parliamentary Questions 7 April 2011, E-002106/2011. Available: ⟨http://www.europarl.europa.eu/sides/getAllAnswers.do?reference=E-2011-002106&language=EN⟩ (accessed 13.01.14).

Dels, J., Cunha, M., Manai, C., Sabugosa-Madeira, B., Silva, M., 2011. Association of financial or professional conflict of interest to research outcomes on health risks or nutritional assessment studies of genetically modified products. Food Policy 36, 197–203.

Dietrich, D., von Aulock, S., Marquardt, H.W.J., Blaauboer, B.J., Dekant, W., Kehrer, J., et al., 2013. Open letter to the European Commission. scientifically unfounded precaution drives European Commission's recommendations on EDC regulation, while defying common sense, well-established science, and risk assessment principles. Arch. Toxicol. 87, 1739–1741.

Domingo, J.L., Bordonaba, J.G., 2011. A literature review on the safety assessment of genetically modified plants. Environ. Int. 37, 734–742.

EC (European Commission), 2013. Flash Eurobarometer 361. Chemicals Report. Available: ⟨http://ec.europa.eu/public_opinion/flash/fl_361_en.pdf⟩ (accessed 30.04.14).

EChA (European Chemicals Agency), 2010. Practical Guide 2. How to Report Weight of Evidence. Helsinki, Finland. Available: ⟨http://echa.europa.eu/documents/10162/13655/pg_report_weight_of_evidence_en.pdf⟩ (accessed 30.04.14).

EChA (European Chemicals Agency), 2011. Chapter R.4. Evaluation of Available Information. In: Guidance on Information Requirements and Chemical Safety Assessment. Helsinki, Finland. Available: ⟨http://echa.europa.eu/documents/10162/13643/information_requirements_r4_en.pdf⟩ (accessed 30.04.14).

EChA (European Chemicals Agency), 2014. Registered Substances – Chemical Substances Search. Available: ⟨http://echa.europa.eu/web/guest/information-on-chemicals/registered-substances⟩ (accessed 25.06.13).

EEB (European Environmental Bureau), ClientEarth, 2012. Identifying the Bottlenecks in REACh Implementation – The Role of EChA in REACh's Failing Implementation. Available: ⟨http://www.eeb.org/EEB/?LinkServID=53B19853-5056-B741-DB6B33B4D1318340⟩ (accessed 14.01.14).

EFSA (European Food Safety Authority), 2010. Application of Systematic Review Methodology to Food and Feed Safety Assessments to Support Decision Making. EFSA Guidance for Those Carrying out Systematic Reviews. EFSA J 8:637. Available: ⟨http://www.efsa.europa.eu/en/efsajournal/pub/1637.htm⟩ (accessed 30.04.14).

EFSA (European Food Safety Authority), 2012a. Europe Needs a Stronger EFSA and a Stronger Risk Assessment Community. Available: ⟨http://www.efsa.europa.eu/en/press/news/121116a.htm⟩ (accessed 30.04.14).

EFSA (European Food Safety Authority), 2012b. When Science Met Society Video. Available: ⟨http://www.youtube.com/watch?v=qDs4FTDLerE&rel=0⟩ (accessed 30.04.14).

EHN (Environmental Health News), 2013. Scientists Critical of EU Chemical Policy have Industry Ties. Available: ⟨http://www.environmentalhealthnews.org/ehs/news/2013/eu-conflict⟩ (accessed 30.04.14).

Eriksson, P., Fischer, C., Wallin, M., Jakobsson, E., Fredriksson, A., 2006. Impaired behaviour, learning and memory, in adult mice neonatally exposed to hexabromocyclododecane (HBCDD). Environ. Toxicol. Pharmacol. 21, 317–322.

Evans, I., Thornton, H., Chalmers, I., Glasziou, P., 2011. Testing Treatments. Better Research for Better Healthcare, 2nd ed. Pinter & Martin Ltd., London (Available: ⟨http://www.worldcat.org/oclc/759000841⟩ (accessed 30.04.14).

Fagin, D., Lavelle, M., 2002. Center for Public Integrity. Toxic Deception. How the Chemical Industry Manipulates Science, Bends the Law & Endangers Your Health, 3rd Ed. Common Courage Press, Monroe, Maine, USA.

Gift, J.S., Caldwell, J.C., Jinot, J., Evans, M.V., Cote, I., Vandenberg, J.J., 2013. Scientific considerations for evaluating cancer bioassays conducted by the Ramazzini Institute. Environ. Health Perspect. 121, 1253–1263.

Gioiosa, L., Fissore, E., Ghirardelli, G., Parmigiani, S., Palanza, P., 2007. Developmental exposure to low-dose estrogenic endocrine disruptors alters sex differences in exploration and emotional responses in mice. Horm. Behav. 52, 307–316.

Gore, A.C., Balthazart, J., Bilde, D., Carpenter, D.O., Crews, D., Czernichow, P., et al., 2013. Policy decisions on endocrine disruptors should be based on science across disciplines. A response to Dietrich et al. Endocrinology 154, 3957–3960.

Green, S., Higgins, J.P.T., Alderson, P., Clarke, M., Mulrow, C.D., Oxman, A.D., 2008. Introduction. In: Higgins, J.P.T., Green, S. (Eds.), Cochrane Handbook for Systematic Reviews of Interventions. John Wiley & Sons, Chichester, UK (Chapter 1).

Hanson, M., Gluckman, P., 2011. Developmental origins of non-communicable disease. Population and public health implications. Am. J. Clin. Nutr. 94 (6 Suppl.), 1754S–1758S.

Hardisty, J.F., 1985. Factors influencing laboratory animal spontaneous tumor profiles. Toxicol. Pathol. 13, 95–104.

Hartung, T., 2009. Food for thought… on evidence-based toxicology. ATLEX 26, 75–82.

Haseman, J.K., 1984. Statistical issues in the design, analysis and interpretation of animal carcinogenicity studies. Environ. Health Perspect. 58, 385–392.

Hayes, T.B., 2004. There is no denying this. Defusing the confusion about atrazine. BioScience 54, 1138–1149.

Hoshi, N., Ohtsuka, T., 2009. Adult rats exposed to low-doses of di-n-butyl phthalate during gestation exhibit decreased grooming behavior. Bull. Environ. Contam. Toxicol. 83, 62–66.

Huff, J., 2002. Chemicals studied and evaluated in long-term carcinogenesis bioassays by both the Ramazzini Foundation and the National Toxicology Program. Ann. N. Y. Acad. Sci. 982, 208–220.

Issaro, C., Samir, H., Zohra, H., Monia, Z., Hassen, B.C., 2009. Toxic responses to deltamethrin (DM) low doses on gonads, sex hormones and lipoperoxidation in male rats following subcutaneous treatments. J. Toxicol. Sci. 34, 663–670.

Johnson, M.D., Kenney, N., Stoica, A., Hilakivi-Clarke, L., Singh, B., et al., 2003. Cadmium mimics the in vivo effects of estrogen in the uterus and mammary gland. Nat. Med. 9, 1081–1084.

Klaassen, C.D., Amdur, M.O., Doull, J., 2013. Casarett & Doull's Toxicology – The Basic Science of Poisons, 8th edition McGraw-Hill, New York, NY.

Klimisch, H.-J., Andreae, M., Tillmann, U., 1997. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. Regul. Toxicol. Pharmacol. 25, 1–5.

Koshland Jr., D., 1998. The era of pathway quantification. Science 280, 852–853.

Kozul-Horvath, C.D., Zandbergen, F., Jackson, B.P., Enelow, R.I., Hamilton, J.W., 2012. Effects of low-dose drinking water arsenic on mouse fetal and postnatal growth and development. PLoS One 7, e38249.

Krauth, D., Woodruff, T.J., Bero, L., 2013. Instruments for assessing risk of bias and other methodological criteria of published animal studies. A systematic review. Environ. Health Perspect. 121, 985–992.

Kuriyama, S.N., Wanner, A., Fidalgo-Neto, A.A., Talsness, C.E., Koerner, W., Chahoud, I., 2007. Developmental exposure to low-dose PBDE-99. Tissue distribution and thyroid hormone levels. Toxicology 242, 80–90.

Lesser, L.I., Ebbeling, C.B., Goozner, M., Wypij, D., Ludwig, D.S., 2007. Relationship between funding source and conclusion among nutrition-related scientific articles. PLoS Med. 4, e5.

Li, X., Pham, H.T., Janesick, A.S., Blumberg, B., 2012. Triflumizole is an obesogen in mice that acts through peroxisome proliferator activated receptor gamma (PPARγ). Environ. Health Perspect. 120, 1720–1726.

Macon, M.B., Villanueva, I.R., Tatum-Gibbs, K., Zehr, R.D., Strynar, M.J., Stanko, J.P., et al., 2011. Prenatal perfluorooctanoic acid exposure in CD-1 mice. Low-dose developmental effects and internal dosimetry. Toxicol. Sci. 122, 134–145.

Makris, S.L., 2011. Current assessment of the effects of environmental chemicals on the mammary gland in guideline rodent studies by the U.S. Environmental Protection Agency (U.S. EPA), Organization for Economic Co-operation and Development (OECD), and National Toxicology Program (NTP). Environ. Health Perspect. 119, 1047–1052.

Maronpot, R.R., Flake, G., Huff, J., 2004. Relevance of animal carcinogenesis findings to human cancer predictions and prevention. Toxicol. Pathol. 32 (Suppl. 1), 40–48.

Meador, J.P., Sommers, F.C., Cooper, K.A., Yanagida, G., 2011. Tributyltin and the obesogen metabolic syndrome in a salmonid. Environ. Res. 111, 50–56.

Melnick, R., Lucier, G., Wolfe, M., Hall, R., Stancel, G., Prins, G., et al., 2002. Summary of the National Toxicology Program's report of the endocrine disruptors low-dose peer review. Environ. Health Perspect. 110, 427–431.

Muñoz-de-Toro, M., Markey, C.M., Wadia, P.R., Luque, E.H., Rubin, B.S., Sonnenschein, C., Soto, A.M., 2005. Perinatal exposure to bisphenol-A alters peripubertal mammary gland development in mice. Endocrinology 146, 4138–4147.

Myers, J.P., vom Saal, F.S., Akingbemi, B.T., Arizono, K., Belcher, S., Colborn, T., et al., 2009. Why public health agencies cannot depend on good laboratory practices as a criterion for selecting data. The case of bisphenol A. Environ. Health Perspect. 117, 309–315.

NIEHS (National Institute of Environmental Health), 2012. Low Dose Effects and Non-monotonic Dose Responses for Endocrine Active Chemicals. In: Science to Practice, An International Workshop. Available: ⟨http://www.niehs.nih.gov/about/visiting/events/pastmtg/2012/dert_endocrine/index.cfm⟩ (accessed 30.04.14).

Nachman, K.E., Fox, M., Sheehan, M.C., Burke, T.A., Rodricks, J.V., Woodruff, T.J., 2011. Leveraging epidemiology to improve risk assessment. Open Epidemiol. J. 4, 3–29.

OECD (Organization for Economic Cooperation and Development), 2010. Cutting Costs in Chemicals Management. How OECD Helps Governments and Industry. Environment Directorate, Paris, France.

OECD (Organization for Economic Cooperation and Development), 2012a. Manual for the Assessment of Chemicals. Paris, France. Environment Directorate, Chapter 3. Available: ⟨http://www.oecd.org/env/ehs/risk-assessment/49191960.pdf⟩ (accessed 30.04.14).

OECD (Organization for Economic Cooperation and Development), 2012b. Guidance Document No. 116. On the Conduct & Design of Chronic Toxicity & Carcinogenicity Studies, Supporting Test Guidelines 451, 452 and 453. 2nd ed. Environment Directorate, Paris, France.

OECD (Organization for Economic Cooperation and Development), 2014a. Mutual Acceptance of Data (MAD). Available: ⟨http://www.oecd.org/env/chemicalsafetyandbiosafety/mutualacceptanceofdatamad.htm⟩ (accessed 30.04.14).

OECD (Organization for Economic Cooperation and Development), 2014b. OECD Guidelines for the Testing of Chemicals and Related Documents. Available: ⟨http://www.oecd.org/env/testguidelines⟩ (accessed 30.04.14).

Olson, K.L., Matsumura, F., Boush, G.M., 1980. Behavioral effects on juvenile rats from perinatal exposure to low levels of toxaphene, and its toxic components, toxicant A and toxicant B. Arch. Environ. Contam. Toxicol. 9, 247–257.

Palanza, P., Parmigiani, S., Liu, H., vom Saal, F.S., 1999. Prenatal exposure to low doses of estrogenic chemicals diethylstilbestrol & o,p'-DDT alters aggressive behavior of male & female house mice. Pharmacol. Biochem. Behav. 64, 665–672.

Park, M., Han, J., Ko, J.J., Lee, W.S., Yoon, T.K., Lee, K., Bae, J., 2011. Maternal exposure to fenarimol promotes reproductive performance in mouse offspring. Toxicol. Lett. 205, 241–249.

Pesticide Action Network Europe, 2012. A TTC for PCPs? mais, oui!. 2012. Available: ⟨http://www.pan-europe.info/Resources/Reports/PAN%20-%202012%20-%20SANCO%20scientific%20committees%20on%20TTC%20-%20conflict%20of%20interest.pdf⟩ (accessed 03.09.13).

Qiao, Y., Li, B., Yang, G., Yao, H., Yang, J., Liu, D., Yan, Y., Sigsgaard, T., Yang, X., 2009. Irritant and adjuvant effects of gaseous formaldehyde on the ovalbumin-induced hyper responsiveness and inflammation in a rat model. Inhal. Toxicol. 21, 1200–1207.

Ray, L.B., Gough, N.R., 2002. Orienteering strategies for a signalling maze. Science 296, 632–633.

Richter, C.A., Birnbaum, L.S., Farabollini, F., Newbold, R.R., Rubin, B.S., Talsness, C.E., et al., 2007. In vivo effects of bisphenol A in laboratory rodent studies. Reprod. Toxicol. 24, 199–224.

Rosenberg, D., 2011. Cancer-Causing Chemicals Have More Friends in Congress than You Do. The Switchboard Blog. Available: ⟨http://switchboard.nrdc.org/blogs/drosenberg/cancer-causing_chemicals_have.html⟩ (accessed 30.04.14).

Rudel, R.A., Fenton, S.E., Ackerman, J.M., Euling, S.Y., Makris, S.L., 2011. Environmental exposures and mammary gland development. State of the science, public health implications, and research recommendations. Environ. Health Perspect. 119, 1053–1061.

Saegusa, Y., Fujimoto, H., Woo, G.H., Inoue, K., Takahashi, M., Mitsumori, K., Hirose, M., Nishikawa, A., Shibutani, M., 2009. Developmental toxicity of brominated flame retardants, tetrabromobisphenol A & 1,2,5,6,9,10-hexabromocyclododecane, in rat offspring after maternal exposure from mid-gestation through lactation. Reprod. Toxicol. 28, 56–67.

Sah, S., Loewenstein, G., 2014. Nothing to declare – mandatory and voluntary disclosure leads advisors to avoid conflicts of interest. Psychol. Sci. 25, 575–584.

Schneider, K., 1983. Faking it: the case against Industrial Bio-Test Laboratories. Amicus Spring, 14–26.

Schug, T.T., Abagyan, R., Blumberg, B., Collins, T.J., Crews, D., DeFur, P.L., et al., 2013. Designing endocrine disruption out of the next generation of chemicals. Green Chem. 15, 181–198.

Soffritti, M., Belpoggi, F., Lambertin, L., Lauriola, M., Padovani, M., Maltoni, C., 2002. Results of long-term experimental studies on the carcinogenicity of formaldehyde and acetaldehyde in rats. Ann. N. Y. Acad. Sci. 982, 87–105.

Soffritti, M., Belpoggi, F., Esposti, D.D., Falcioni, L., Bua, L., 2006. Consequences of exposure to carcinogens beginning during developmental life. Basic Clin. Pharmacol. Toxicol. 102, 118–124.
Sterne, J.A.C., Egger, M., Moher, D., 2008. Addressing reporting biases. In: Higgins, J.P.T., Green, S. (Eds.), Cochrane Handbook for Systematic Reviews of Interventions. John Wiley & Sons, Chichester UK (Chapter 10).
Sturtz, N., Jahn, G.A., Deis, R.P., Rettori, V., Duffard, R.O., Evangelista de Duffard, A. M., 2010. Effect of 2,4-dichlorophenoxyacetic acid on milk transfer to the litter and prolactin release in lactating rats. Toxicology 271, 13–20.
Suvorov, A., Takser, L., 2011. Delayed response in the rat frontal lobe transcriptome to perinatal exposure to the flame retardant BDE-47. J. Appl. Toxicol. 31, 477–483.
Swaen, G.M., Meijers, J.M., 1988. Influence of design characteristics on the outcome of retrospective cohort studies. Br. J. Ind. Med. 45, 624–629.
Tilson, H.A., Schroeder, J.C., 2013. Reporting of results from animal studies (Editorial). Environ. Health Perspect. 121, A320–A321.
USCDC (United States Centers for Disease Control and Prevention). National Health and Nutrition Examination Survey (NHANES). Available: ⟨http://www.cdc.gov/nchs/nhanes.htm⟩ (accessed 30.04.14).
USEPA (United States Environmental Protection Agency), 2010. Dioxin Homepage. Available: ⟨http://cfpub.epa.gov/ncea/CFM/nceaQFind.cfm?keyword=Dioxin⟩ (accessed 30.04.14).
USEPA (United States Environmental Protection Agency), 2014a. Harmonized Test Guidelines. Available: ⟨http://www.epa.gov/ocspp/pubs/frs/home/guidelin.htm⟩ (accessed 30.04.14).
USEPA (United States Environmental Protection Agency), 2014b. The Integrated Risk Information System Homepage. Available: ⟨http://www.epa.gov/iris/index.html⟩ (accessed 30.04.14).
USEPA (United States Environmental Protection Agency), 2014c. Integrated Risk Assessment Information System (IRIS). Cadmium. Available: ⟨http://www.epa.gov/iris/subst/0141.htm⟩ (accessed 25.06.13).
USEPA (United States Environmental Protection Agency). 2014d. Integrated Risk Assessment Information System (IRIS). Arsenic. Available: ⟨http://www.epa.gov/iris/subst/0278.htm⟩ (accessed 25.06.13).
USEPA (United States Environmental Protection Agency), 2014e. Integrated Risk Assessment Information System (IRIS). Formaldehyde. Available: ⟨http://www.epa.gov/iris/subst/0419.htm⟩ (accessed 25.06.13).
USEPA (United States Environmental protection Agency), 2014f. Integrated Risk Assessment Information System (IRIS). Haloxyfop methyl. Available: ⟨http://www.epa.gov/iris/subst/0467.htm⟩ (accessed 16.05.14).
USEPA (United States Environmental Protection Agency), 1999. Draft Guidance on Developing Robust Summaries. Available: ⟨http://www.epa.gov/HPV/pubs/general/robsumgd.htm⟩ (accessed 30.04.14).
USFDA (United States Food and Drug Administration), 2014. Inspections, Compliance, Enforcement, and Criminal Investigations. Available: ⟨http://www.fda.gov/ICECI/EnforcementActions/BioresearchMonitoring/default.htm⟩ (accessed 30.04.14).
USNAS (United States National Academy of Sciences), 2009. Science and Decisions, Advancing Risk Assessment. National Academies Press, Washington, DC. Available: ⟨http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=202175⟩ (accessed 30.07.13).
USNRC (United States National Research Council), 1983. Risk Assessment in the Federal Government: Understanding the Process. National Academy Press, Washington DC. Available: ⟨http://www.nap.edu/openbook.php?isbn=0309033497⟩ (accessed 16.05.14).
USNRC (United States National Research Council), 2014a. State-of-the-Science Evaluation of Nonmonotonic Dose–Response Relationships as They Apply to Endocrine Disruptors. National Academes Press, Washington DC. Available: ⟨http://www.nap.edu/catalog.php?record_id=18608⟩ (accessed 02.05.14).
USNRC (United States National Research Council). 2014b. Review of EPA's Integrated Risk Information System (IRIS) Process. National Academies Press, Washington DC. Available: ⟨http://www.nap.edu/catalog.php?record_id=18764⟩ (accessed 06.05.14).
USNTP (United States National Toxicology Program), 2013a. Draft OHAT Approach for Systematic Review and Evidence Integration for Literature-Based Health Assessments. Available: ⟨http://ntp.niehs.nih.gov/NTP/OHAT/EvaluationProcess/DraftOHATApproach_February2013.pdf⟩ (accessed 30.04.14).
USNTP (United States National Toxicology Program), 2013b. Webinar on the Assessment of Data Quality in Animal Studies. Available: ⟨http://ntp.niehs.nih.gov/go/38752⟩ (accessed 30.04.14).
Valkusz, Z., Nagyéri, G., Radács, M., Ocskó, T., Hausinger, P., László, M., et al., 2011. Further analysis of behavioral and endocrine consequences of chronic exposure of male Wistar rats to subtoxic doses of endocrine disruptor chlorobenzenes. Physiol. Behav. 103, 421–430.
Vandenberg, L.N., 2014. Non-monotonic dose responses in studies of endocrine disrupting chemicals: bisphenol a as a case study. Dose Response 12, 259–276.
Vesterinen, H.M., Johnson, P.I., Koustas, E., Lam, J., Sutton, P., Woodruff, T.J., 2013. In support of EHP's proposal to adopt the ARRIVE guidelines (letter). Environ. Health Perspect. 121, A325.
Vos, J.G., De Klerk, A., Krajnc, E.I., Van Loveren, H., Rozing, J., 1990. Immunotoxicity of bis(tri-n-butyltin)oxide in the rat. Effects on thymus-dependent immunity and on nonspecific resistance following long-term exposure in young versus aged rats. Toxicol. Appl. Pharmacol. 105, 144–155.
Vosges, M., Braguer, J.C., Combarnous, Y., 2008. Long-term exposure of male rats to low-dose ethinylestradiol (EE2) in drinking water. Effects on ponderal growth and on litter size of their progeny. Reprod. Toxicol. 25, 161–168.
vom Saal, F., Hunt, P. Opinion 2012. FDA's Decision on BPA Exposes Catch 22. Environmental Health News. Available: ⟨http://www.environmentalhealthnews.org/ehs/news/2012/op-ed-the-fdas-catch-22⟩ (accessed 30.04.14).
vom Saal, F.S., Hughes, C., 2005. An extensive new literature concerning low-dose effects of Bisphenol A shows the need for a new risk assessment. Environ. Health Perspect. 113, 926–933.
Walker, A.I., Stevenson, D.E., Robinson, J., Thorpe, E., Roberts, M., 1969. The toxicology and pharmacodynamics of dieldrin (HEOD): two-year oral exposures of rats and dogs. Toxicol. Appl. Pharmacol. 15, 345–373.
Wignall, J.A., Shapiro, A.J., Wright, F.A., Woodruff, T.J., Chiu, W.A., Guyton, K.Z., Rusyn, I., 2014. Standardizing benchmark dose calculations to improve science-based decisions in human health assessments. Environ. Health Perspect. 122, 499–505.
Woodruff, T.J., Sutton, P., 2010. Pulling back the curtain. improving reviews in environmental health. Environ. Health Perspect. 118, a326–a327.
Woodruff, T.J., Sutton, P., 2011. An evidence-based medicine methodology to bridge the gap between clinical and environmental health sciences. Health Aff. (Millwood) 30, 31–37.
Yu, K.O., Narayanan, L., Matute, D.R., Godfrey, R.J., Todd, P.N., Sterner, T.R., et al., 2002. The pharmacokinetics of perchlorate & its effect on hypothalamus–pituitary–thyroid axis in male rat. Toxicol. Appl. Pharmacol. 182, 148–159.
Yu, P.L., Lin, H.W., Wang, S.W., Wang, P.S., 2011. Effects of nonylphenol on the production of progesterone on the rats granulosa cells. J. Cell. Biochem. 112, 2627–2636.
Zinser, D.E., Bolcic-Jankovic, D., Clarridge, B., Blumenthal, D., Campbell, E.G., 2009. Participation of academic scientists in relationships with industry. Health Aff. (Millwood) 28, 1814–1825.